The Registry Data Warehouse in the European Reference Network for Rare Respiratory Diseases - Background, Conception and Implementation

Schaaf, Jannik; Chalmers, James; Omran, Heymut; Pennekamp, Petra; Sitbon, Olivier; Wagner, Thomas O. F.

Published in:
German Medical Data Sciences

DOI:
10.3233/SHTI210049

Publication date:
2021

German Medical Data Sciences: Bringing Data to Life
R. Röhrig et al. (Eds.)

doi:10.3233/SHTI210049

The Registry Data Warehouse in the European Reference Network for Rare Respiratory Diseases – Background, Conception and Implementation

Jannik SCHAAF[a,1], James CHALMERS[b], Heymut OMRAN[c], Petra PENNEKAMP[c], Olivier SITBON[d], Thomas O.F. WAGNER[e], Abilio REIS[f], Dennis KADIOGLU[a] and Holger STORF[a]

[a] *Medical Informatics Group, University Hospital Frankfurt, Frankfurt, Germany*
[b] *Scottish Centre for Respiratory Research, University of Dundee, Dundee, United Kingdom*
[c] *Department of General Pediatrics, University of Münster, Münster, Germany*
[d] *Centre de Référence de l'Hypertension Pulmonaire, Hôpital Bicêtre, AP-HP, Université Paris-Saclay, France*
[e] *Frankfurt Reference Center for Rare Diseases, University Hospital Frankfurt, Frankfurt, Germany*
[f] *Unidade de Doença Vascular Pulmonar, Centro Hospitalar Universitário do Porto, Porto, Portugal*

Abstract. Rare lung diseases affect 1.5-3 million people in Europe while causing bad prognosis or early deaths for patients. The European Reference Network for Respiratory Diseases (ERN-Lung) is a patient centric network, funded by the European Union (EU). The aims of ERN-LUNG is to increase healthcare and research regarding rare respiratory diseases. An initial need for cross-border healthcare and research is the use of registries and databases. A typical problem in registries for RDs is the data exchange, since the registries use different kind of data with different types or descriptions. Therefore, ERN-Lung decided to create a new Registry Data-Warehouse (RDW) where different existing registries are connected to enable cross-border healthcare within ERN-Lung. This work facilitates the aims, conception and implementation for the RDW, while considering a semantic interoperability approach. We created a common dataset (CDS) to have a common descriptions of respiratory diseases patients within the ERN registries. We further developed the RDW based on Open Source Registry System for Rare Diseases (OSSE), which includes a Metadata Repository with the Samply.MDR to unique describe data for the minimal dataset. Within the RDW, data from existing registries is not stored in a central database. The RDW uses the approach of the "Decentral Search" and can send requests to the connected registries, whereas only aggregated data is returned about how many patients with specific characteristics are available. However, further work is needed to connect the different existing registries to the RDW and to perform first studies.

[1] Corresponding Author, Jannik Schaaf, Medical Informatics Group, University Hospital Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany; E-mail: schaaf@med.uni-frankfurt.de.

Keywords. Rare Diseases, Data Integration, Patient Registry

1. Introduction

Rare lung diseases affect about 1.5 up to 3 million people in Europe. Some of these diseases belong to mucociliary clearance (MCC) disorders [1]. Such diseases are Cystic Fibrosis (CF), Primary Ciliary Dyskinesia (PCD) and non-CF bronchiectasis (non-CF BE). These complex lung diseases require multidisciplinary care due to the complexity, as these diseases can affect multiple organ systems. Furthermore, other rare lung diseases like Pulmonary Arterial Hypertension (PAH) are causing bad prognosis for patients [2]. Thus, early diagnosis and access to specialists in this field are necessary in order to improve outcomes [3].

The European Reference Network for Respiratory diseases (ERN-Lung) is a patient-centric network of healthcare providers and patient organisations. ERN-Lung is one of 24 European Reference Networks, funded by the European Union (EU). These ERNs include 23 rare or low prevalence complex diseases, conditions and disease-groups, including more than 300 hospitals [4]. ERN-Lung currently includes 60 healthcare providers in 12 different countries, addressing rare respiratory diseases [3]. ERN-Lung is divided in core networks, which include for instance research and clinical trials, registries and biobanks, cross-border care and continued medical education [3]. However, to support clinical research for RDs, patient's registries and databases are needed. The small number of patients affected by RDs make registries even more important than in the normal population or in common diseases as they enable cross-border gathering of data to achieve sufficient sample size for any purpose in clinical research and or public health issues [5,6]. A typical problem in registries for RDs is the data exchange [7]. Due to the disease specific and heterogeneity of data, registries do not provide any semantic interoperability to exchange data [5,8]. A system is considered as semantic interoperable if systems can exchange data with an unambiguous, shared meaning [9].

To enable data exchange as the basis for clinical research in ERN-Lung, the consortia have developed a new Registry Data-Warehouse (RDW) to enable cross-border healthcare through using semantic interoperability, while considering common standards, ethical and data protection regulations. The objectives of this work are to present the aims, a conception and software-implementation of the RDW, as well as an interoperability approach between existing ERN-Lung registries. For reporting our work, we follow the ISCIL Schema (Introduction – State of the art – Concept – Implementation – Lessons learned).

Aims of the Registry Data Warehouse

The aims of the RDW are to allow data exchange as the basis for clinical research in ERN-Lung. Existing and new registries will be brought together in a comprehensive registry landscape with interoperability across possible facets. The RDW provides appropriate software systems and components to achieve these goals. Clinical care registries will be able to exchange data with epidemiologic data registries as well as with healthcare utilization or health economic data registries.

To enable the RDW in ERN-Lung, we conducted a conception and implementation phase. In both stages, we considered needs and ideas from all registries stakeholders, like clinicians and technical experts. In the conception phase, we involved all stakeholders in

the development of a common dataset (CDS). This record is stored in all registries and in the RDW and thus enables a uniform description of the patients. The data was described with a Metadata Repository (MDR) (shown in section 3 and 4). In the implementation phase, we implemented the RDW based on several software-components described in section 4.

2. State of the art

Currently, different registries in the ERN-Lung exist (shown in Table 1, which has been collected with a questionnaire by the registry operators). For a prototypical implementation of the RDW, we have chosen to create a semantic interoperability concept to connect the following registries: A registry for CF (European Cystic Fibrosis Society Patient Registry, https://www.ecfs.eu/ecfspr), PCD (International Primary Ciliary Dyskinesia (PCD) Registry , https://www.pcdregistry.eu), non-CF BE (European Bronchiectasis Registry , https://www.bronchiectasis.eu/registry) and PAH (French Pulmonary Hypertension Registry). As starting with these four registries, we will create a semantic interoperability concept, with the goal to involve all ERN-Lung registries in the future.

Table 1. Results of the questionnaire for all visualization methods of the mockups

Registry	Productive since	Amount patients	Software solution
European Bronchiectasis Registry	2015	18000	Own development
International Primary Ciliary Dyskinesia (PCD) Registry (PCDregistry)	2015	1176	Own development
European Cystic Fibrosis Society Patient Registry	2008	48000	Own development
French Pulmonary Hypertension Registry (FPHR)	2007	13,800	PAHTool

3. Concept

The RDW include the following three major elements (as shown in Figure 1): The ERN-Lung Registry, ERN-Lung Population Registry and disease specific registries (CF, PCD, non-CF BE and PAH). The ERN-Lung Registry (1) will serve as an internal system for documentation and analysis, whereas data from other registries are connected there. The ERN-Lung Registry includes a CDS. The ERN-Lung Registry is governed by the ERN-Lung Board (2). Another component is the ERN-Lung Population Registry (4) which offers an opportunity for patients to self-register in the registry. Between ERN-Lung Registry and ERN-Lung Population Registry, bidirectional semantic interoperability is aimed with a CDS (3).

Most patients can only be reached by patient organizations, experts or specialized care centres. We want to attract as many patients affected by rare pulmonary diseases to get in contact with the ERN and its services. By this option we hope to get in contact with new patients who self-register in the ERN-Lung Population Registry for closing the gap of knowledge concerning the number of patients affected. This will help them to actively contribute to RD research and thus the ERN-Lung Population Registry can

contribute to patient empowerment. The ERN Patient Portal will allow patient recorded outcomes once the patient is registered with the system (5). The RDW will allow bidirectional semantic interoperability (6) between the ERN-Lung Registry, the Population Registry and diseases specific registries (CF, PCD, non-CF BE and PAH) (7). Interoperability between existing registries with a CDS is a prerequisite (8). If data is to be given to a third party, an anonymised copy of the relevant data is generated (9). This is a one-direction export functionality (10). Data sharing with the RDW (data copy export only) upon request might be asked for by the European Commission's Joint Research Centre (JRC) of the European Commission in Ispra, Italy (11). Furthermore, to make data semantically interoperable, an ERN-wide Meta Data Repository (MDR) will be established. A MDR is a dictionary of data-elements, describing the clinical data in a registry. The functionality will be increased by a Registry & Metadata Repository (RMDR), which stores information about which registries are connected and which are using which kind of data (13). To further support collaborations between researchers, a central ERN ID-Management System (14), described in section 4.3, will be provided to create unique pseudonyms in an untraceable manner based on the patients' identifying data. With the help of the included 'record linkage' functionality, redundancies of patient entries can be avoided.

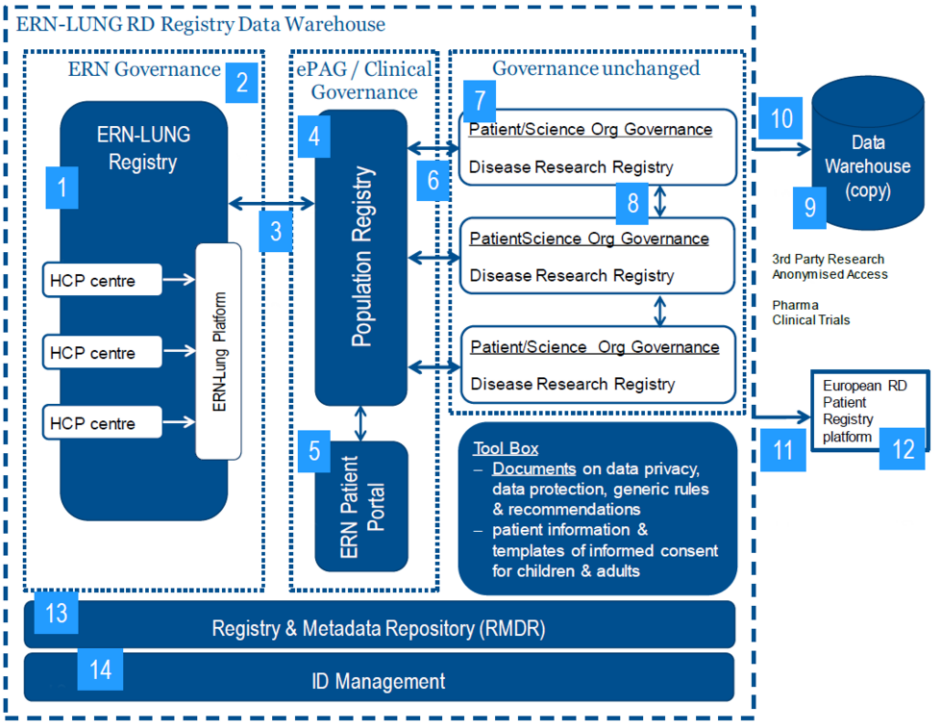

Figure 1. Structure of the Registry Data-Warehouse

Minimal dataset of ERN-Lung registries

The ERN-Lung Registry and Population Registry will contains a set of common data elements for RDs as well as specific data elements customized for ERN purposes like

cross-border consultation. Both registries are making use of the same minimal data set as all other registry applications (CF, PCD, non-CF BE and PAH). The minimal dataset is based on the recommendation of the JRC of the European Commission. The ERN-Lung Registry, as well as all other registries (CF, PCD, non-CF BE and PAH) will implement these dataset to ensure semantic interoperability between the registries. The datasets contains 16 elements of the JRC minimal dataset, collecting basic information about the patients, e.g. diagnosis, diagnosis-history and age at onset. This data is expanded by six further data elements for ERN-Lung administrative reasons, e.g. collection of information needed for cross-border consultation [10].

4. Implementation

In this section, we describe the software-components that were considered for the implementation of the RDW. We will introduce the Open Source Registry System for Rare Diseases (OSSE) as a software-framework for the ERN-Lung Registry as well as the ERN-Lung Population Registry. Furthermore, we show the Mainzelliste as a software for ID- Management and the MDR to describe the medical data in the registries.

4.1 Using OSSE for the ERN-Lung Data Warehouse

As stated in the previous section, two registries in context of the RDW were developed. The ERN-Lung Registry and the ERN-Lung Population Registry. Both registries were developed with OSSE, which allows patient organizations, researchers and physicians to create patient registries based on an open-source software. OSSE is currently used in various international and national registries, e.g. in the European lipodystrophy registry [11]. OSSE was funded by the German Federal Ministry of Health in context of the National Plan for Rare Diseases in 2013. In order to reach semantic interoperability between registries, OSSE considered at an early stage the conception of a high level of interoperability. An important aspect for this is provided with the metadata concept [5]. It is a central component of OSSE and will be described in the next section.

There are several approaches to merge data from our local registries into the ERN-Lung Registry. First, the data can be uploaded to a central registry or remain decentralized at their locations. The first approach can be critical, since data protection laws did not allow to duplicate or upload the data into another system. OSSE supports the approach of decentralised data storage. This means that the data remains in the local registries. OSSE provides a component called “search broker” to connect the registries between each other, called “decentral search”. A request, defined as a description of metadata can be send to one local registry, to check, if a sufficient amount of patients with specific characteristics are available. The request needs an approval of the data owner of the local registry, who will check the request. Outside the registry, a data exchange between both parties is possible [5]. For all registries in ERN-Lung, we aim to connect the registries with the decentral approach.

4.2 Meta Data Repository & Registry & Metadata Repository (RMDR)

Metadata is defined as data that describes the actual data concerning their structure and representation. This concept is used for the creation and documentation of data in OSSE registries and establishes the basis for data exchange between the registries. OSSE uses

the software module Samply.MDR (abbreviated as MDR in the following) to describe the data to be stored in a registry. The MDR is a software to support the specification of data elements in a structured, formalized and standardized way. MDR was originally developed in the context of the German Cancer Consortium [12]. The MDR uses the metadata standard ISO 11179 part 3 edition 3, which describes how metadata can be described [13]. As an example for data element definition, we will show the definition of “blood pressure” as follows. A data element called "blood pressure" is created in the MDR and a short description of the collected characteristic is stored. The data element contains two integer values: systolic blood pressure and diastolic blood pressure. It is stored which unit is used for the blood pressure measurement (mmHg). This is necessary, since different medical parameters can be described with different units (e.g. blood glucose with mg/dl or mmol/l). Furthermore, a validity check of each data element can be defined, to consider, that a value is in a certain value range or not. This means for the mentioned example, that a blood pressure value is never negative and therefore always above zero [5]. Since the MDR is connected to the OSSE-registry, electronic case report forms (eCRF) can be created based on the MDR. This enables the data input into the registry to be based on the data elements of the MDR [5]. By using the MDRs in combination with a CDS, semantic interoperability between all ERN-Lung registries can be ensured.

As the MDR only covers metadata about data elements and not about the registries collecting these, our initial approach was to extend the functionality of the MDR. However, the European Rare Diseases Registry Infrastructure (ERDRI) has been developed by the JRC in February 2019. ERDRI makes rare diseases registries searchable and findable. Core elements of ERDRI are the Central Metadata Repository (ERDRI.MDR) and the European Directory of Registries (ERDRI.DOR). The ERDRI.MDR contains all metadata used by participating registries with their specific designations and definitions. With ERDRI.MDR metadata items can manually be entered using a graphical user interface. While establishing a new registry, a user can search for already existing data element specifications to select them instead of creating new data elements. The ERDRI.DOR gives an overview of all existing registries, including characteristics of these registries. In order to avoid parallel structures at European Level, we decided to use the ERDRI.MDR and the ERDRI.DOR for the RDW [14].

To use the ERDRI-Infrastructure in the RDW, adaptations to various components of the registry software OSSE were necessary. The OSSE Electronic Data Capture and the Collaboration Client, which is part of the search broker, were adapted connect to the ERDRI-Infrastructure.

4.3 ID-Management

Whenever patient data are being used for research, identity management is of crucial importance for reasons of data protection. Identity Management has to separate the identifying data (I-DAT) of the patient from the medical data (M-DAT) and to hide these identifying data by means of a pseudonym or by irreversibly delink these data types by means of anonymization. Identity Management fulfils the central data protection requirement: by means of pseudonymization in combination with a trusted third party, an accidental or intentional misuse of the data is made significantly more difficult. By separated storage of the different data types, an informational separation of powers is being achieved.

A special aspect of medical registries is that often several institutions cooperate with their own pseudonymization schemes. Therefore, the Identity Management must also be able to manage various pseudonyms and to merge them if necessary. Since pseudonyms preserve the linkage between medical data and identity data, the data remains personal in the legal sense, unlike with anonymization. The storage and processing of the pseudonyms must therefore be covered by a corresponding declaration of informed consent by the patient.

To enable pseudonymization of patients, the OSSE architecture includes the local ID management/pseudonymization service for each national or regional OSSE registry. The ID management software used in OSSE is called Mainzelliste [5,15]. The software Mainzelliste was released as Open Source by the Medical Center of Johannes Gutenberg University Mainz and is already being used in various other projects.

5. Lessons learned (Discussion)

The European Commission recommends to enable registries to improve research and healthcare in rare diseases [16]. The ERN-Lung consortia follows this recommendation and have developed the ERN-Lung RDW.

In this work we have shown the aims, a conception and software-implementation of the RDW, as well as a semantic interoperability approach between ERN-Lung registries. The use of OSSE, a metadata concept and a CDS, enables semantic interoperability between the RDW and the local registries. By using the CDS of the JRC of the European Commission, as well as the connection to ERDRI, makes the RDW usable and available on a European level. However, other ERNs are also developing similar structures within their networks. Using a CDS among all ERNs brings opportunity for interoperability. However, a standard is needed to enable data-exchange between the ERNs and their registries. A possibility could be the FAIR approach. This approach makes it possible to make data of registries **F**indable, **A**ccessible, **I**nteroperable and **R**eachable for research and healthcare. This concept is also considered and implemented in OSSE with the MDR and the decentral search [17].

However, our work is currently limited to the conception and implementation of the ERN-Lung RDW. We have not described further legal or data-protection aspects like patient consent, which are not necessary for the implementation. However, patients are currently recruited by the registry operators and a data protection concept based on the General Data Protection Regulation (GDPR) is created and discussed. However, OSSE is complaint with the comprehensive data protection handbook for medical research network of the TMF e.V., a technology and methods platform and umbrella organization for networked medical research [18].

6. Conclusion

The RDW provides a possibility to connect RDs registries with a focus on respiratory diseases and enables semantic interoperability between the registries to increase research and healthcare. Further work is needed to connect the local registries via the OSSE-bridgehead and to consider patient consent in the RDW. Currently, patient consent is limited to the local registries. We hope the RDW and their data will allow a detail portrait of patients with different rare respiratory diseases and enables an improvement of patients affected with those diseases. By applying the Population Registry within the RDW, we invite all physicians and researcher to participate in the registry.

References

[1] McCarthy C, Gallego B, Trapnell B.C., McCormack F.X., Epidemiology of Rare Lung Diseases: The Challenges and Opportunities to Improve Research and Knowledge, *Adv. Exp. Med. Biol.* **1031** (2017), 419–442.

[2] Riou M, Pizzimenti M, Enache I, Charloux A, Canuet M, Andres E et al., Skeletal and Respiratory Muscle Dysfunctions in Pulmonary Arterial Hypertension, *J. Clin. Med.* **9** (2020).

[3] European Reference Network Respiratory Diseases (ERN-Lung), (2020). https://ern-lung.eu/ (accessed February 10, 2020).

[4] Héon-Klin V, European Reference networks for rare diseases: what is the conceptual framework?, *Orphanet J. Rare Dis.* **12** (2017), 137–137.

[5] Storf H, Schaaf J, Kadioglu D, Goebel J, Wagner T.O.F., Ueckert F, Registries for rare diseases : OSSE - An open-source framework for technical implementation, *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. **60** (2017), 523–531.

[6] Forrest C.B, Bartek R.J., Rubinstein Y, Groft S.C., The case for a global rare-diseases registry, *The Lancet*. **377** (2011), 1057–1059.

[7] Blumenthal S, Improving Interoperability between Registries and EHRs, *AMIA Jt. Summits Transl. Sci.* (2018), 20–25..

[8] Sernadela P, Gonzalez-Castro L, Carta C, van der Horst E, Lopes P, Kaliyaperumal R et al., Linked Registries: Connecting Rare Diseases Patient Registries through a Semantic Web Layer, *BioMed Res. Int.* **(**2017) ,8327980.

[9] Arvanitis T.N., Semantic interoperability in healthcare, *Stud. Health Technol. Inform.* **202** (2014), 5–8.

[10] European Comission Joint Research Centre, SET OF COMMON DATA ELEMENTS FOR RARE DISEASES REGISTRATION, (2020). http://www.erare.eu/sites/default/files/SetCommonData-EU%20RD%20Platform_CDS%20_final.pdf (accessed February 10, 2020).

[11] von Schnurbein J, Adams C, Akinci B, Ceccarini G, D'Apice M.R., Gambineri A et al., European lipodystrophy registry: background and structure, *Orphanet J. Rare Dis.* **15** (2020), 17.

[12] Kadioglu D, Breil B, Lablans M, Mate S, Schlue D, Serve H et al., Samply MDR - A Metadata Repository and Its Application in Various Research Networks, *Stud Health Technol Inform*. (2018), 50–54.

[13] International Organization for Standardization, ISO/IEC 11179, Information technology – Metadata registries, (2013).

[14] European Comission, European Rare Disease Registry Infrastructure (ERDRI), (2019). https://eu-rd-platform.jrc.ec.europa.eu/erdri-description (accessed February 10, 2020).

[15] Institut für Medizinische Biometrie, Epidemologie und Informatik (IMBEI), Die Mainzelliste-Pseudonymisierung und Identitätsmanagement, (2019). http://www.unimedizin-mainz.de/imbei/informatik/ag-verbundforschung/mainzelliste.html (accessed February 21, 2019).

[16] AHRQ, Registries for evaluating patient outcomes: a user's guide, in: Methods Eff. Health Care, 3rd ed., Rockville, 2014.

[17] Schaaf J, Kadioglu D, Goebel J, Behrendt C.A., Roos M, van Enckevort D et al., OSSE Goes FAIR - Implementation of the FAIR Data Principles for an Open-Source Registry for Rare Diseases, *Stud. Health Technol. Inform.* **253** (2018), 209–213.

[18] OSSE - Overall Concept. https://www.osse-register.de/en/overall-concept/ (accessed July 14, 2020).